

500/441 28 JUN 2004  
(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION  
EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété  
Intellectuelle  
Bureau international



(43) Date de la publication internationale  
10 juillet 2003 (10.07.2003)

PCT

(10) Numéro de publication internationale  
**WO 03/056455 A1**

(51) Classification internationale des brevets<sup>7</sup> : G06F 17/30

(71) Déposant (pour tous les États désignés sauf US) :  
**FRANCE TELECOM** [FR/FR]; 6, place d'Alleray,  
F-75015 Paris (FR).

(21) Numéro de la demande internationale :  
PCT/FR02/04549

(22) Date de dépôt international :  
24 décembre 2002 (24.12.2002)

(72) Inventeurs; et  
(75) Inventeurs/Déposants (pour US seulement) : **RODET,**  
**Xavier** [FR/FR]; 5, avenue de la Dame Blanche, F-94120  
Fontenay-sous-Bois (FR). **WORMS, Laurent** [FR/FR];  
36, rue de Villacoublay, F-78140 Velizy-Villacoublay  
(FR). **PEETERS, Geoffroy** [BE/FR]; 12, allée Vivaldi,  
F-75012 Paris (FR).

(25) Langue de dépôt : français

(26) Langue de publication : français

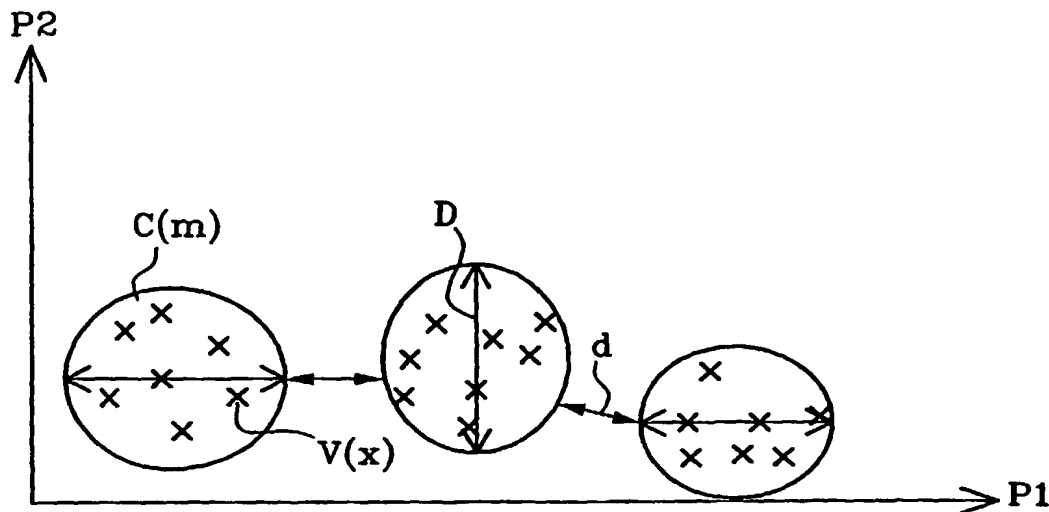
(30) Données relatives à la priorité :  
01/16949 27 décembre 2001 (27.12.2001) FR

(74) Mandataires : **SOMNIER, Jean-Louis** etc.; Cabinet Bal-  
lot, 122, rue Edouard Vaillant, F-92593 Levallois-Perret  
Cedex (FR).

[Suite sur la page suivante]

(54) Title: METHOD FOR CHARACTERIZING A SOUND SIGNAL

(54) Titre : PROCÉDE DE CARACTERISATION D'UN SIGNAL SONORE



WO 03/056455 A1  
(57) Abstract: The invention concerns a method for characterizing in accordance with specific parameters, a sound signal  $x(t)$  varying in time  $t$  in different frequency bands  $k$ , and referenced  $x(k, t)$ . It consists in storing the signal  $x(t)$ , calculating and storing the energy  $E(k, t)$  of said signal  $x(k, t)$  for each of said bands  $k$ ,  $k$  varying from 1 to  $K$  and in accordance with a time window  $h(t)$  of duration  $2N$ , and in a second step, calculating the energy variation and the signal phase  $E(k, t)$  in  $J$  frequency bands, the  $J$  values referenced  $F(j, k, t)$  and  $\phi(j, k, t)$  thus obtained constituting the specific parameters of an extract of duration  $2N$  of the sound signal  $x(t)$  and in repeating said calculation at every  $S$  time interval.

(57) Abrégé : L'invention concerne un procédé de caractérisation selon des paramètres spécifiques, d'un signal sonore  $x(t)$  évoluant selon le temps  $t$  dans différentes bandes de fréquences  $k$  et alors noté  $x(k, t)$ . Il consiste à mémoriser le signal  $x(t)$ , à calculer et mémoriser l'énergie  $E(k, t)$  dudit signal  $x(k, t)$  pour chacune desdites bandes de fréquences  $k$ ,  $k$  variant de 1 à  $k$  et selon une fenêtre temporelle  $h(t)$  d'une durée  $2N$ , et dans un second

[Suite sur la page suivante]



(81) États désignés (*national*) : AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

européen (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR),  
brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Publiée :**

— avec rapport de recherche internationale

(84) États désignés (*régional*) : brevet ARIPO (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), brevet eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet

*En ce qui concerne les codes à deux lettres et autres abréviations, se référer aux "Notes explicatives relatives aux codes et abréviations" figurant au début de chaque numéro ordinaire de la Gazette du PCT.*

**PROCEDE DE CARACTERISATION D'UN SIGNAL SONORE**

L'invention concerne un procédé de caractérisation selon des paramètres spécifiques, d'un signal sonore évoluant selon le temps dans différentes bandes de fréquences.

5        Le domaine de l'invention est celui de la reconnaissance de signaux sonores appliquée en particulier à l'identification d'œuvres musicales utilisées sans autorisation.

10        En effet, le développement des techniques de numérisation et du multimédia a provoqué une augmentation considérable de telles utilisations frauduleuses. Il en résulte une nouvelle difficulté pour les organismes chargés de collecter les droits d'auteur, puisqu'il faut être en mesure d'identifier  
15        ces utilisations en particulier sur les réseaux numériques interactifs tel Internet, afin d'effectuer de façon satisfaisante la perception et la répartition des rémunérations dues aux auteurs de ces œuvres musicales.

20        Dans la suite, pour ne pas se limiter aux œuvres musicales, on considérera plus généralement un signal sonore.

25        Le but de la présente invention est donc de constituer une base de données de signaux sonores, chaque signal sonore étant caractérisé par une empreinte de sorte que, étant donné un signal sonore inconnu que l'on caractérise de la même façon, on puisse effectuer une recherche et une comparaison

rapides de l'empreinte de ce signal inconnu avec l'ensemble des empreintes de la base de données.

L'empreinte est constituée de paramètres spécifiques déterminés de la façon suivante.

5        On décompose dans un premier temps le signal sonore dont l'amplitude  $x(t)$  varie avec le temps  $t$ , selon différentes bandes de fréquences  $k$  :  $x(k,t)$  est l'amplitude du signal sonore filtré dans la bande de fréquence  $k$  et représenté figure 1a).

10        On calcule, comme représenté figure 1c) l'énergie à court terme  $E(k,t)$  de ce signal sonore filtré à l'aide d'une fenêtre  $h(t)$  représentée figure 1b), ayant un support de  $2N$  secondes. On réitère ce calcul en faisant glisser cette fenêtre toutes les  $S$  secondes.

15        Ces valeurs  $E(k,t)$  constituent des paramètres spécifiques d'un extrait de  $2N$  secondes du signal sonore  $x(k,t)$  dans la bande de fréquences  $k$ .

20        On peut obtenir d'autres paramètres en calculant pour différentes bandes de fréquences  $j$ , l'énergie de  $E(k,t)$  à l'aide d'une fenêtre  $h'(t)$  représentée figure 2b), ayant un support de  $2N'$  secondes ; on réitère ce calcul en faisant glisser cette fenêtre toutes les  $S'$  secondes : on obtient  $F(j,k,t)$  représenté figure 2c). On normalise les valeurs  $F(j,k,t)$  par rapport à leur  
25        maximum de façon à les rendre indépendantes de l'amplitude du signal sonore.

Ainsi normalisées, ces valeurs constituent des paramètres spécifiques d'un extrait de  $2N'$  secondes du signal sonore  $x(k,t)$  dans la bande de fréquences  $k$ .

30        On peut également calculer la phase de  $E(k,t)$  pour différentes bandes de fréquences  $j$  : on obtient

$P(j,k,t)$ . On normalise les valeurs  $P(j,k,t)$  par rapport à une valeur de référence  $P(1,j,t)$  et on obtient alors d'autres paramètres spécifiques d'un extrait de  $2N'$  secondes du signal sonore.

5 D'autres paramètres peuvent être ajoutés tels que la valeur moyenne de l'énergie  $E(k,t)$ .

L'invention a pour objet un procédé de caractérisation selon des paramètres spécifiques, d'un  
10 signal sonore  $x(t)$  évoluant selon le temps  $t$  pendant une durée  $D$  dans différentes bandes de fréquences  $k$  et alors noté  $x(k,t)$ , principalement caractérisé en ce qu'il consiste à mémoriser le signal  $x(t)$ , à calculer l'énergie  $E(k,t)$  dudit signal  $x(k,t)$  pour chacune  
15 desdites bandes de fréquences  $k$ ,  $k$  variant de 1 à  $K$  et selon une fenêtre temporelle  $h(t)$  d'une durée  $2N$ , mémoriser les valeurs de l'énergie  $E(k,t)$  obtenues, ces valeurs constituant les paramètres spécifiques d'un extrait d'une durée  $2N$  du signal sonore  $x(t)$  et à  
20 réitérer ce calcul à intervalles réguliers pour obtenir l'ensemble des paramètres spécifiques pour la durée  $D$  du signal sonore  $x(t)$ .

Il consiste en outre à calculer et mémoriser l'énergie  $F(k,j,t)$  de  $E(k,t)$  pour des bandes de  
25 fréquences  $j$ ,  $j$  variant de 1 à  $J$ , selon une fenêtre temporelle  $h'(t)$  d'une durée  $2N'$ , les  $J \times K$  valeurs de l'énergie  $F(j,k,t)$  obtenues constituant les paramètres spécifiques d'un extrait d'une durée  $2N'$  du signal sonore  $x(t)$  et à réitérer ce calcul à intervalles

réguliers pour obtenir l'ensemble des paramètres spécifiques pour la durée  $D$  du signal sonore  $x(t)$ .

Il consiste éventuellement à calculer la phase  $P(j,k,t)$  de l'énergie  $E(k,t)$  pour des bandes de fréquences  $j$ ,  $j$  variant de 1 à  $J$ ,  $j$  différent de  $k$ , et à inclure les valeurs de la phase  $P(j,k,t)$  obtenues parmi les paramètres spécifiques du signal sonore  $x(t)$ .

Il peut aussi consister à calculer pour chaque bande de fréquence  $j$ , la valeur moyenne de l'énergie  $E(k,t)$  sur  $2N'$  secondes, à réitérer ce calcul à intervalles réguliers pour obtenir l'ensemble des paramètres spécifiques pour la durée  $D$  du signal sonore  $x(t)$  et à inclure les valeurs moyennes obtenues parmi les paramètres spécifiques du signal sonore  $x(t)$ .

Selon une caractéristique, il consiste à considérer les paramètres spécifiques d'un signal sonore  $x(t)$  comme les composantes d'un vecteur représentatif de  $x(t)$ , à positionner les vecteurs dans un espace à autant de dimensions que de paramètres, à définir des classes regroupant les vecteurs les plus proches et à enregistrer lesdites classes.

Les classes présentant des distances inter-classes et des distances intra-classes, le procédé consiste avantageusement à sélectionner parmi les paramètres spécifiques, les paramètres permettant d'obtenir des distances inter-classes relativement grandes devant les distances intra-classes et à enregistrer les paramètres sélectionnés.

L'invention concerne également un dispositif d'identification d'un signal sonore, caractérisé en ce qu'il comprend un serveur de base de données comprenant

des moyens pour la mise en œuvre du procédé de caractérisation d'un signal sonore selon des paramètres spécifiques, tel que décrit précédemment et des moyens de recherche dudit signal sonore dans la base de données.

De préférence, les moyens de recherche comprennent des moyens de reconnaissance directe de la classe à laquelle ledit signal sonore appartient et des moyens de recherche de la classe par comparaison des paramètres spécifiques du signal sonore inconnu avec ceux de la base de donnée, la classe étant choisie, par exemple, par la méthode des plus proches voisins.

D'autres particularités et avantages de l'invention apparaîtront clairement à la lecture de la description faite à titre d'exemple non limitatif et en regard des dessins annexés sur lesquels :

les figures 1a), 1b) et 1c) représentent respectivement des courbes schématiques de variation d'un signal sonore  $x(k_i, t)$  filtré dans une bande de fréquences  $k_i$ , d'une fenêtre de Hamming  $h(t)$  et de l'énergie  $E(k_i, t)$  à court terme du signal  $x(k_i, t)$ ,

les figures 2a), 2b) et 2c) représentent respectivement des courbes schématiques de variation de l'énergie  $E(k_i, t)$  pour la bande de fréquence  $k_i$ , d'une fenêtre de Hamming  $h'(t)$  et de l'énergie  $F(j_m, k_i, t)$  de  $E(k_i, t)$  pour la bande de fréquences  $j_m$ ,

la figure 3 illustre schématiquement un ensemble de vecteurs  $V[x(t)]$  constituant l'empreinte d'un signal  $x(k, t)$ ,

la figure 4 illustre schématiquement le stockage d'empreintes,

la figure 5 représente une classification des signaux sonores selon deux paramètres,

5 la figure 6 illustre une méthode de recherche d'un signal sonore par la méthode du plus proche voisin.

la figure 7 représente schématiquement un serveur de base de données de stockage des empreintes de signaux sonores.

10

Les signaux sonores que l'on traite selon ce procédé de caractérisation sont des signaux sonores enregistrés, notamment sur des disques compacts (Compact Disk en anglais).

15

On va considérer par la suite que le signal sonore  $x(t)$  est un signal numérique échantillonné à une fréquence d'échantillonnage  $f_e$ , par exemple 11025 Hz correspondant au quart de la fréquence d'échantillonnage courante des disques compacts qui est de 44100 Hz.

20

On peut cependant caractériser un signal sonore analogique : il faut au préalable le convertir en un signal numérique au moyen d'un convertisseur analogique-numérique.

25

Le signal sonore  $x(k,t)$  représenté figure 1a) pour  $k=k_i$  est donc un signal numérique échantillonné à la fréquence  $f_e$  et obtenu après un filtrage dans une bande de fréquences  $k_i$ . Chaque valeur de ce signal numérique échantillonné est codée par exemple sur 16 bits. Les bandes de fréquences sont des bandes du spectre audible variant d'environ 20 Hz à 20kHz et découpé en  $K$  ( $k$

30



varie de 1 à K) bandes de fréquences, K=127 par exemple.

L'énergie à court terme  $E(k,t)$  représentée figure 1c) pour  $k=k_i$ , est calculée sur une fenêtre  $h(t)$  de 2N secondes, par exemple une fenêtre de Hamming ayant un support d'environ 23 ms représentée figure 1b).

$E(k,t)$  est le carré du module d'une transformée du signal sonore échantillonné  $x(t)$  dans le plan temps-fréquence ou dans le plan temps-échelle. Parmi les transformées utilisables, figurent la transformée de Fourier, la transformée en cosinus, la transformée de Hartley et la transformée en ondelettes. Un banc de filtres passe-bandes réalise également une telle transformée. La transformée de Fourier à court terme permet une représentation temps-fréquence adaptée à l'analyse des signaux musicaux. Ainsi l'énergie  $E(k,t)$  s'écrit :

$$E(k,t) = \left| \sum_{n=-N}^{n=N} x(t+n/f_e).h(n/f_e).e^{-4i\pi kn/N} \right|^2$$

avec  $i$  tel que  $i^2 = -1$

On fait glisser la fenêtre sur le signal sonore toutes les S secondes, par exemple toutes les 10 ms.  $E(k,t)$  sera ainsi échantillonné toutes les 10 ms : on aura  $E(k,t_0)$  puis  $E(k,t_1)$  avec  $t_1=t_0+10$  ms etc.

On code ainsi toutes les S secondes le signal sonore  $x(t)$  par un vecteur à K composantes  $E(k,t)$ , chacune de ces composantes codant l'énergie de 23 ms du signal sonore  $x(t)$  dans K bandes de fréquences.

On obtient d'autres paramètres en reproduisant en quelque sorte les calculs précédents et en les appliquant cette fois à  $E(k,t)$  comme représenté figure 2a) à figure 2c).

5 On filtre l'énergie  $E(k,t)$  dans  $J$  différentes bandes de fréquence :  $E(j,k,t)$  est l'énergie  $E(k,t)$  filtrée dans la bande de fréquences  $j$ ,  $j$  variant de 1 à  $J$  avec par exemple  $J=51$ .

10 On calcule alors  $F(j,k,t)$  représentée figure 2c) pour  $k=k_i$  et  $j=j_m$ , sur une fenêtre  $h'(t)$  de  $2N'$  secondes, par exemple une fenêtre de Hamming ayant un support de 10 s. Ainsi, on peut écrire, avec  $i$  tel que  $i^2 = -1$  :

$$F(j,k,t) = \left| \sum_{n=-N'}^{n=N'} E(k,t+n/f_e) \cdot h'(n/f_e) \cdot e^{-4i\pi j n/N'} \right|^2$$

15

Dans notre exemple, toutes les secondes ( $S'=1$ ), le signal sonore  $x(t)$  est codé par  $127 \times 51$  paramètres  $F(j,k,t)$ , chaque réel  $F(j,k,t)$  représentant l'énergie de dix secondes ( $2N'=10$ ) du signal d'énergie  $E(k,t)$  dans la bande de fréquence  $j$ .

20 De manière à rendre  $F(j,k,t)$  indépendant de l'amplitude du signal qui peut être plus ou moins fort, on rapporte ces valeurs à une valeur de référence, en l'occurrence la valeur maximale de  $F_M(j,k,t)$  pour tous les  $k$  et  $j$  considérés. On obtient ainsi  $K \times J$  paramètres  
25  $F(j,k,t)/F_M(j,k,t)$ .

On calcule en outre toutes les  $2N'$  secondes, la phase de l'énergie  $E(k,t)$  dans chacune des bandes de fréquences  $j$ :  $P(j,k,t)$ .

- 5 Pour cela, on calcule l'argument de la transformée de Fourier de  $E(k,t)$  dans chacune des bandes de fréquence  $j$  :

$$P(j,k,t) = \text{Arg} \left| \sum_{n=-N'}^{n=N'} E(k,t+n/f_c) \cdot h'(n/f_c) \cdot e^{-4i\pi j n/N'} \right|$$

- 10 Comme précédemment, on rapporte ces valeurs à une valeur de référence, en l'occurrence la valeur de  $P(j,k,t)$  pour la seconde bande de fréquences ( $j=1$ ) considérée car la référence temporelle de l'échantillon est inconnue : on ne connaît pas l'origine des temps.

- 15 Pour cela, on calcule les phases rapportées  $\varphi(j,k,t)$  par les formules suivantes :

$$\varphi(1,k,t) = P(1,k,t)$$

$$\varphi(j,k,t) = P(j,k,t) - P(1,k,t) \cdot \frac{f(k)}{f(1)}, \text{ pour } k > 1.$$

- 20 où les  $f(k)$  sont les fréquences centrales des canaux  $k$ .  
On obtient ainsi  $K \times J$  paramètres correspondant aux valeurs de la phase rapportée  $\varphi(j,k,t)$ .

- 25 On peut également considérer d'autres paramètres notamment, les valeurs moyennes de l'énergie  $E(k,t)$  sur  $2N'$  secondes et ce pour chaque bande de fréquences  $j$  :  $E(j,k,t)$ .

L'ensemble de ces paramètres normalisés définissent à intervalles réguliers, une empreinte que l'on peut considérer comme un vecteur  $V(x(t))$ . L'ensemble des paramètres normalisés par exemple  
5  $F(j,k,t)/F_M$  et  $P(j,k,t)-P(j,1,t)$  définissent toutes les  $S'$  secondes, une empreinte que l'on peut considérer comme un vecteur  $V(x(t))$  à  $2 \times K \times J$  dimensions ( $2 \times 127 \times 51$  soit environ 13000 dans notre exemple), une dimension par paramètre, chaque vecteur caractérisant un extrait  
10 de  $2N'$  secondes du signal sonore  $x(t)$ , 10 secondes dans notre exemple.

On réitère cette caractérisation toutes les  $S'$  secondes, toutes les secondes par exemple ( $S'=1$ ).

Comme représenté figure 3, un signal  $x(t)$  durant  $T$   
15 secondes est finalement caractérisé par  $L$  vecteurs  $V$ ,  $L$  étant environ égal à  $T/S'$ .

Pour un signal sonore durant 10 mn soit 600s, on obtient 600 vecteurs c'est-à-dire  $600 \times 2 \times J \times K$  paramètres.

Ces vecteurs sont mémorisés dans la zone de  
20 stockage 10 d'une base de données hébergée sur un serveur ou sur un disque compact. On a représenté figure 4 l'ensemble des vecteurs  $V$  d'un signal ou d'une œuvre  $A$  par  $V_A$ , de même  $V_B$  pour une œuvre  $B$ , etc.

Il est souhaitable de réduire le nombre de  
25 composantes de ces vecteurs c'est-à-dire le nombre de paramètres de façon à obtenir un vecteur ou une empreinte de taille plus réduite en vue de son stockage dans la base de données. De plus, lorsqu'il s'agira de comparer l'empreinte d'un signal sonore inconnu avec  
30 celles de la base de données, il sera souhaitable que

le nombre de paramètres à comparer soit réduit pour que cette recherche soit effectuée rapidement.

Or ces paramètres n'apportent pas tous la même quantité d'information, certaines pouvant être  
5 redondantes voire inutiles. C'est pourquoi on sélectionne les paramètres les plus pertinents parmi tous les paramètres, par un calcul d'information mutuelle présenté notamment dans la publication PROC. ICASSP'99, Phoenix, Arizona, USA, Mars 1999 H.YANG, S.VAN VUUREN, H.HERMANSKY, « Relevancy of Time-Frequency Features for Phonetic Classification Measured  
10 by Mutual Information ». On limite ainsi  $K$  à  $K_1$  et  $J$  à  $J_1$ .

On va présenter une méthode de sélection de ces  
15 paramètres.

Chacune des empreintes de ces signaux sonores c'est-à-dire chacun de ces vecteurs est classé dans un espace  $R$  à  $N$  dimensions,  $N$  étant le nombre de composantes des vecteurs. Pour des raisons de  
20 simplification, un exemple de classification pour des vecteurs à 2 dimensions  $P_1$  et  $P_2$  est représenté figure 5.

On définit des classes  $C(m)$  regroupant les vecteurs par proximité,  $m$  variant de 1 à  $M$ . On peut par  
25 exemple décider qu'une classe correspond à une œuvre musicale : dans ce cas  $M$  est le nombre d'œuvres musicales mémorisées dans la base de données.

Il résulte du calcul de l'information mutuelle entre ces classes  $C(m)$  et les paramètres, que la  
30 pertinence des paramètres est liée aux distances inter et intra classes : des paramètres pertinents

garantissent des distances inter-classes d relativement grandes comparées aux distances intra-classes D.

En ne retenant que les paramètres pertinents, on définit ainsi  $K_1$  et  $J_1$ .

5        On peut par exemple considérer cinq ( $K_1=5$ ) bandes de fréquences respectivement centrées sur 344 Hz, 430 Hz, 516 Hz, 608 Hz et 689 Hz.

Des essais ont été effectués en prenant  $J_1 = 3$ .

10       Les classes  $C(m)$  sont alors constituées à partir des vecteurs  $V_q(x)$  ne comportant plus que  $2 \times K_1 \times J_1$  composantes.

On va donner un exemple pour  $K_1=5$  et  $J_1=3$ , de la taille mémoire d'une base de données contenant 1000 heures de musique et en considérant comme paramètres  
15        $E(k,t)$  et  $F(j,k,t)$ , chacun de ces paramètres étant codé sur 4 octets.

Les paramètres  $E(k,t)$  calculés toutes les 10 ms occupent  $1000 \times 3600 \times 100 \times 5 \times 4$  octets soit environ 7 Giga octets.

20       Les paramètres  $F(j,k,t)$  calculés toutes les secondes occupent  $1000 \times 3600 \times 3 \times 5 \times 4$  octets soit environ 200 Méga octets.

Ces paramètres sont associés aux références des signaux sonores ; si l'on considère que les références  
25       contiennent 100 caractères codés chacun sur un octet, ces références occupent  $1000 \times 10 \times 100$  octets soit environ 1 Méga octet.

Une telle base de données occupera finalement environ 7 Giga octets.

Lorsqu'on souhaite identifier un signal sonore inconnu, on en établit d'abord l'empreinte, référencée  $V(x_{inc})$  à la figure 6, comme décrit précédemment, sachant que le signal sonore inconnu peut être une œuvre musicale complète ou un extrait de cette œuvre.

La recherche de la classe de cette empreinte dans la base de données consiste alors selon une méthode classique illustrée figure 6, à comparer les paramètres de cette empreinte  $V(x_{inc})$  à ceux des empreintes de la base de donnée. Les empreintes les plus proches, dites les plus proches voisins, définissent la classe de la façon suivante : la classe est celle de la majorité des plus proches voisins.

Un serveur de base de données 1 est schématiquement représenté figure 7. Il comprend une zone de stockage 10 des données de la base dans laquelle sont mémorisées les empreintes des signaux sonores assorties de leurs références. Il comprend en outre une mémoire 11 dans laquelle sont stockés les programmes de caractérisation et de recherche précédemment décrits, un processeur 12 assorti de mémoires de travail pour mettre en œuvre ces programmes. Il comprend bien sûr une interface d'entrée-sortie 13 et un bus 14 reliant ces divers éléments entre eux.

Lorsqu'il s'agit de rentrer dans la base de données 1, de nouveaux signaux sonores, l'interface 13 reçoit le signal  $x(t)$  assorti de ses références ; s'il s'agit seulement d'un signal inconnu à identifier, l'interface 13 reçoit seulement le signal  $x(t)$  inconnu.

En sortie, l'interface 13 fournit une réponse à la  
recherche d'un signal inconnu. Cette réponse est  
négative si le signal inconnu n'existe pas dans la zone  
de stockage 10 ; si le signal a été identifié, la  
5 réponse comporte les références du signal identifié.

10

15



**REVENDEICATIONS**

1. Procédé de caractérisation selon des paramètres spécifiques, d'un signal sonore  $x(t)$  évoluant selon le temps  $t$  pendant une durée  $D$  dans différentes bandes de fréquences  $k$  et alors noté  $x(k,t)$ , caractérisé en ce  
5 qu'il consiste à mémoriser le signal  $x(t)$ , à calculer l'énergie  $E(k,t)$  dudit signal  $x(k,t)$  pour chacune desdites bandes de fréquences  $k$ ,  $k$  variant de 1 à  $K$  et selon une fenêtre temporelle  $h(t)$  d'une durée  $2N$ , mémoriser les valeurs de l'énergie  $E(k,t)$  obtenues, ces  
10 valeurs constituant les paramètres spécifiques d'un extrait d'une durée  $2N$  du signal sonore  $x(t)$  et à réitérer ce calcul à intervalles réguliers pour obtenir l'ensemble des paramètres spécifiques pour la durée  $D$  du signal sonore  $x(t)$ .

15

2. Procédé selon la revendication précédente, caractérisé en ce qu'il consiste à calculer et mémoriser l'énergie  $F(k,j,t)$  de  $E(k,t)$  pour des bandes de fréquences  $j$ ,  $j$  variant de 1 à  $J$ , selon une fenêtre  
20 temporelle  $h'(t)$  d'une durée  $2N'$ , les  $J \times K$  valeurs de l'énergie  $F(j,k,t)$  obtenues constituant les paramètres spécifiques d'un extrait d'une durée  $2N'$  du signal sonore  $x(t)$  et à réitérer ce calcul à intervalles réguliers pour obtenir l'ensemble des paramètres  
25 spécifiques pour la durée  $D$  du signal sonore  $x(t)$ .

3. Procédé selon l'une des revendications précédentes, caractérisé en ce qu'il consiste à calculer la phase rapportée  $\varphi(j,k,t)$  de l'énergie  $E(k,t)$  pour des bandes de fréquences  $j$ ,  $j$  variant de 1 à  $J$ , et à inclure les valeurs de la phase rapportée  $\varphi(j,k,t)$  obtenues parmi les paramètres spécifiques du signal sonore  $x(t)$ .

4. Procédé selon l'une des revendications précédentes, caractérisé en ce qu'il consiste à calculer pour chaque bande de fréquence  $j$ , la valeur moyenne de l'énergie  $E(k,t)$  sur  $2N'$  secondes, à réitérer ce calcul à intervalles réguliers pour obtenir l'ensemble des paramètres spécifiques pour la durée  $D$  du signal sonore  $x(t)$  et à inclure les valeurs moyennes obtenues parmi les paramètres spécifiques du signal sonore  $x(t)$ .

5. Procédé selon l'une des revendications précédentes, caractérisé en ce qu'il consiste à considérer les paramètres spécifiques d'un signal sonore  $x(t)$  comme les composantes d'un vecteur représentatif de  $x(t)$ , à positionner les vecteurs dans un espace à autant de dimensions que de paramètres, à définir des classes regroupant les vecteurs les plus proches et à enregistrer lesdites classes.

6. Procédé selon la revendication précédente, caractérisé en ce que les classes présentent des distances inter-classes et des distances intra-classes et en ce qu'il consiste à sélectionner parmi les

paramètres spécifiques, les paramètres permettant d'obtenir des distances inter-classes relativement grandes devant les distances intra-classes et à enregistrer les paramètres sélectionnés.

5

7. Dispositif d'identification d'un signal sonore, caractérisé en ce qu'il comprend un serveur de base de données comprenant des moyens pour la mise en œuvre du procédé de caractérisation d'un signal sonore selon des paramètres spécifiques, selon l'une quelconque des revendications précédentes et des moyens de recherche dudit signal sonore dans la base de données.

8. Dispositif selon la revendication précédente prise en combinaison avec la revendication 5 ou 6, caractérisé en ce que les moyens de recherche comprennent des moyens de reconnaissance de la classe à laquelle ledit signal sonore appartient et des moyens de comparaison, par la méthode du plus proche voisin, des paramètres spécifiques du signal sonore inconnu avec les paramètres spécifiques de la base de données.

25

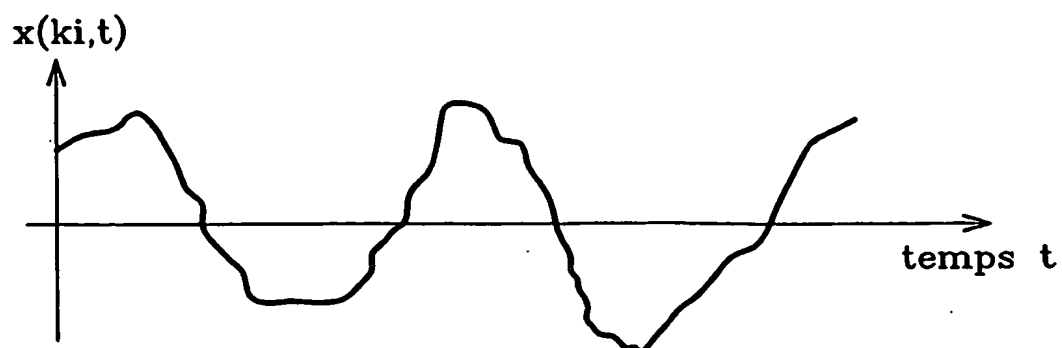


Fig. 1a

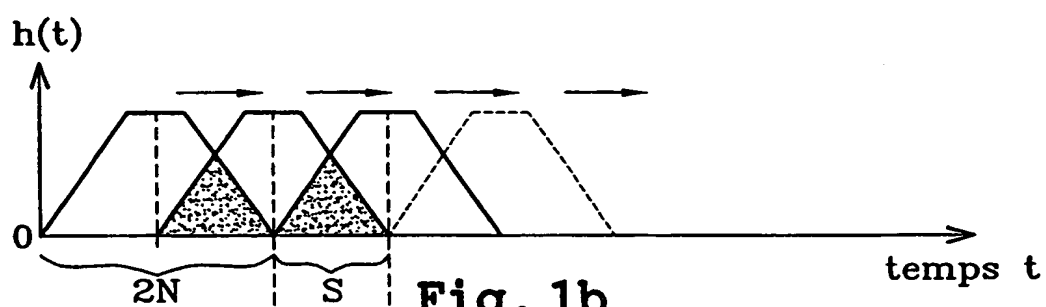


Fig. 1b

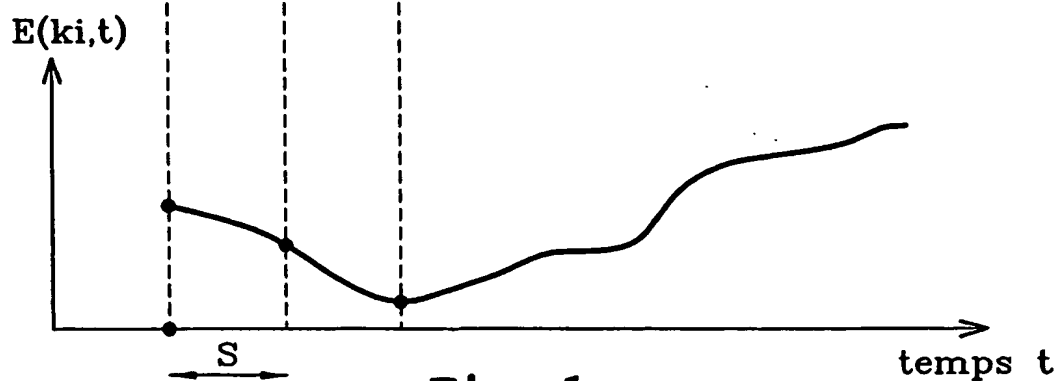


Fig. 1c

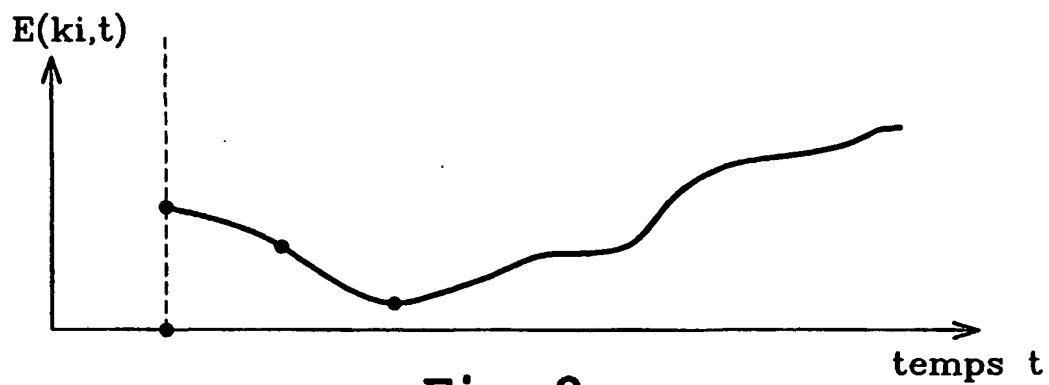


Fig. 2a

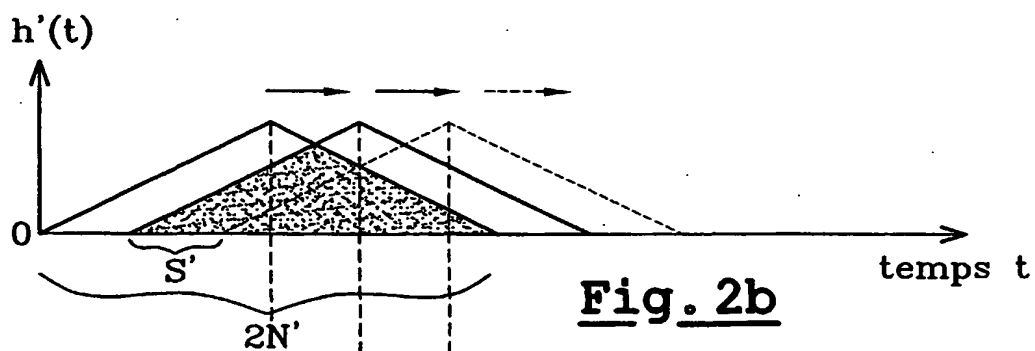


Fig. 2b

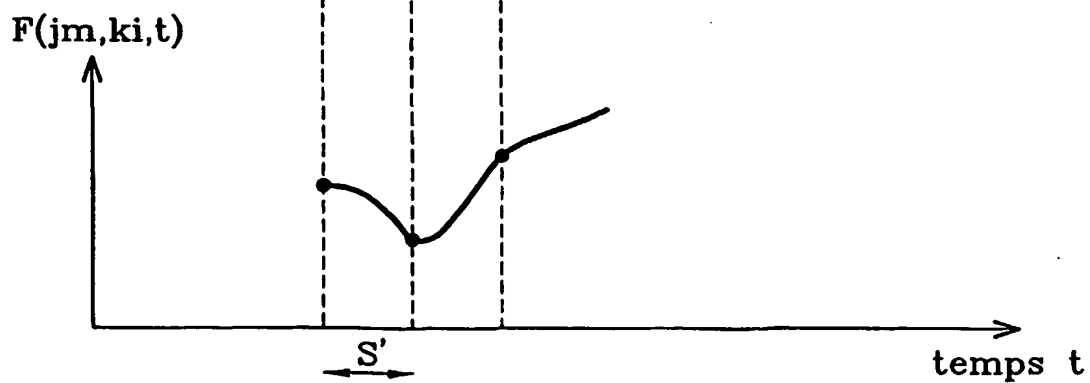
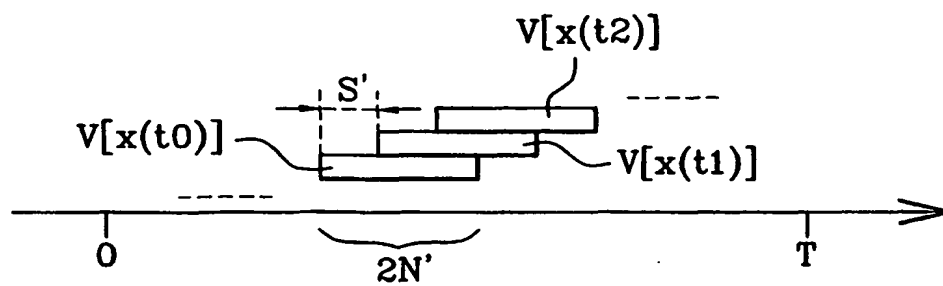
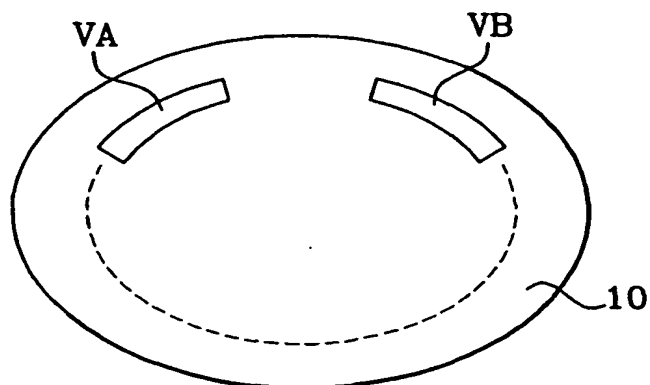
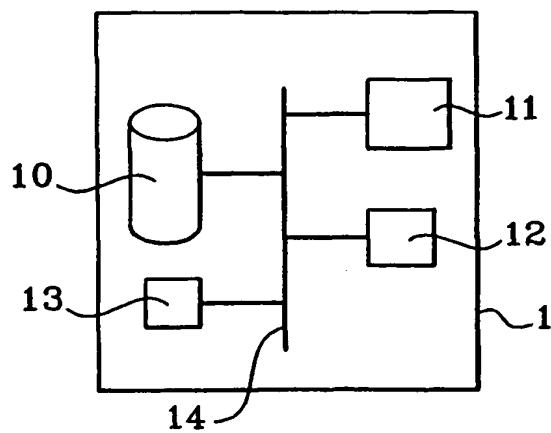
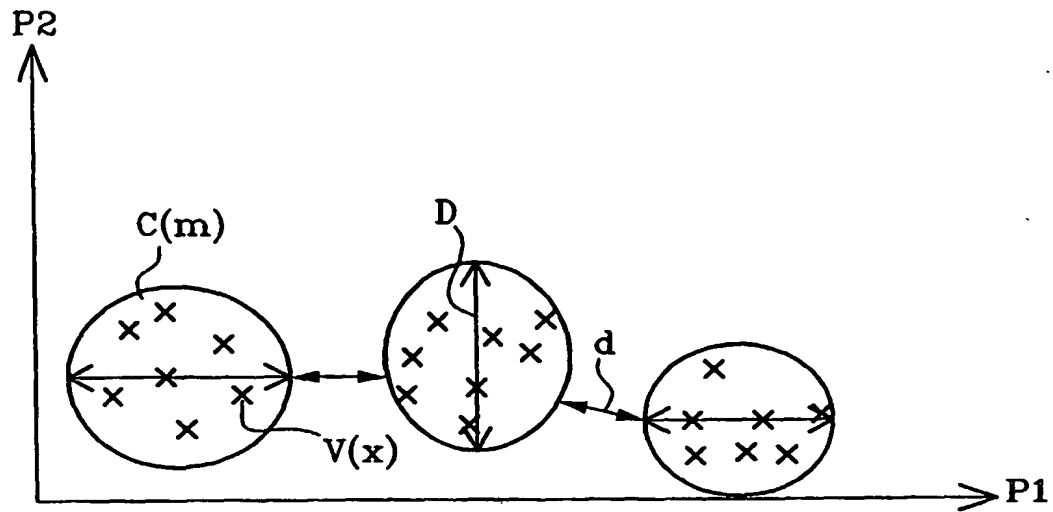
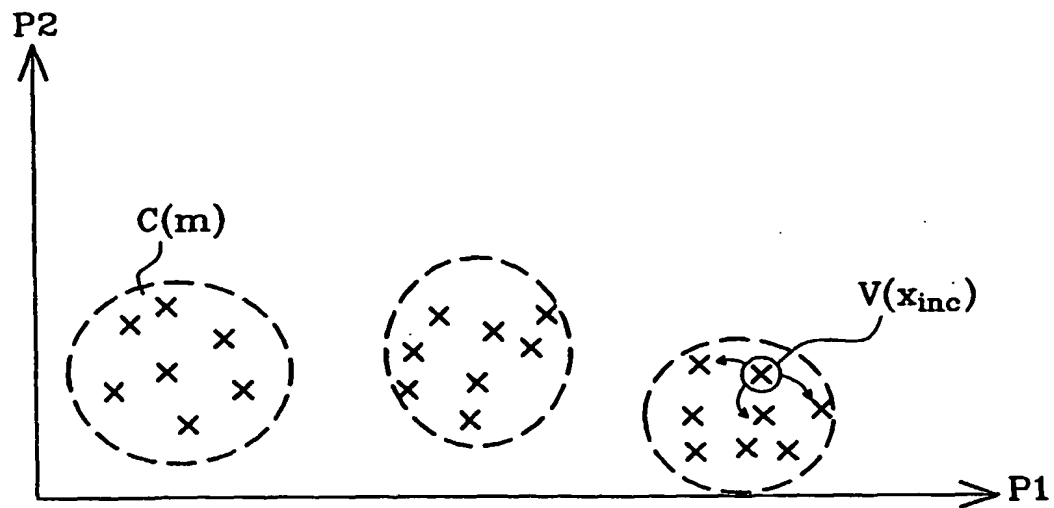


Fig. 2c

**Fig. 3****Fig. 4****Fig. 7**

Fig. 5Fig. 6

## INTERNATIONAL SEARCH REPORT

Int. nat. application No.  
PCT/FR 04549A. CLASSIFICATION OF SUBJECT MATTER  
IPC 7 G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
IPC 7 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 955 592 A (CANON KK) 10 November 1999 (1999-11-10) page 4, line 35 -page 8, line 15 ---	1-8
X	WOLD E ET AL: "Content-based classification, search, and retrieval of audio" IEEE MULTIMEDIA, IEEE COMPUTER SOCIETY, US, vol. 3, no. 3, 1996, pages 27-36, XP002154735 ISSN: 1070-986X the whole document --- -/--	1-8

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

## \* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*G\* document member of the same patent family

Date of the actual completion of the international search

8 April 2003

Date of mailing of the international search report

15/04/2003

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

DE CASTRO PALOM., L



## INTERNATIONAL SEARCH REPORT

Int. nat. Application No

PCT/FR 04549

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5 918 223 A (KEISLAR DOUGLAS F ET AL) 29 June 1999 (1999-06-29) column 3, line 4 -column 3, line 67 column 5, line 21 -column 8, line 65 column 17, line 8 -column 19, line 34 ---	1-8
A	GONZALEZ & K MELIH R: "Content Based Retrieval of Audio" PROCEEDINGS OF THE AUSTRALIAN TELECOMMUNICATION NETWORKS AND APPLICATIONS CONFERENCE, XX, XX, 1996, XP002154738 the whole document -----	1-8

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/FR 04549

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0955592	A	10-11-1999	AU 718035 B2 06-04-2000
		AU 2385499 A	25-11-1999
		EP 0955592 A2	10-11-1999
		JP 2000035796 A	02-02-2000
		US 6201176 B1	13-03-2001
US 5918223	A	29-06-1999	NONE

# RAPPORT DE RECHERCHE INTERNATIONALE

Classification internationale No  
PCT/FR/04549

**A. CLASSEMENT DE L'OBJET DE LA DEMANDE**  
CIB 7 G06F17/30

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

**B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE**

Documentation minimale consultée (système de classification suivi des symboles de classement)  
CIB 7 G06F

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si réalisable, termes de recherche utilisés)  
EPO-Internal, WPI Data

**C. DOCUMENTS CONSIDERES COMME PERTINENTS**

Catégorie *	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	EP 0 955 592 A (CANON KK) 10 novembre 1999 (1999-11-10) page 4, ligne 35 -page 8, ligne 15 ---	1-8
X	WOLD E ET AL: "Content-based classification, search, and retrieval of audio" IEEE MULTIMEDIA, IEEE COMPUTER SOCIETY, US, vol. 3, no. 3, 1996, pages 27-36, XP002154735 ISSN: 1070-986X le document en entier --- -/--	1-8

☒ Voir la suite du cadre C pour la fin de la liste des documents

☒ Les documents de familles de brevets sont indiqués en annexe

\* Catégories spéciales de documents cités:

- \*A\* document définissant l'état général de la technique, non considéré comme particulièrement pertinent
- \*E\* document antérieur, mais publié à la date de dépôt international ou après cette date
- \*L\* document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)
- \*O\* document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens
- \*P\* document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

- \*T\* document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention
- \*X\* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément
- \*Y\* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier
- \*G\* document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

8 avril 2003

Date d'expédition du présent rapport de recherche internationale

15/04/2003

Nom et adresse postale de l'administration chargée de la recherche internationale  
Office Européen des Brevets, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Fonctionnaire autorisé

DE CASTRO PALOM..., L

# RAPPORT DE RECHERCHE INTERNATIONALE

Den Int. No  
PCT/FR 04549

## C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
A	US 5 918 223 A (KEISLAR DOUGLAS F ET AL) 29 juin 1999 (1999-06-29) colonne 3, ligne 4 - colonne 3, ligne 67 colonne 5, ligne 21 - colonne 8, ligne 65 colonne 17, ligne 8 - colonne 19, ligne 34 ---	1-8
A	GONZALEZ & K MELIH R: "Content Based Retrieval of Audio" PROCEEDINGS OF THE AUSTRALIAN TELECOMMUNICATION NETWORKS AND APPLICATIONS CONFERENCE, XX, XX, 1996, XP002154738 le document en entier -----	1-8

# RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres des familles de brevets

De : [redacted] Internationale No

PCT/FR [redacted] 04549

Document brevet cité au rapport de recherche		Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
EP 0955592	A	10-11-1999	AU 718035 B2	06-04-2000
			AU 2385499 A	25-11-1999
			EP 0955592 A2	10-11-1999
			JP 2000035796 A	02-02-2000
			US 6201176 B1	13-03-2001
<hr/>				
US 5918223	A	29-06-1999	AUCUN	
<hr/>				